nominal (e.g., single instance) amount of data. The social behavior recognition system **10** may take any pair of observed behaviors and determine whether the two behaviors match. The social behavior recognition system **10** may use various cameras **12** that are not borne by any individuals (e.g., located remote from the people **14**). The cameras **12** may be used to capture and analyze non-verbal cues (e.g., emotional affect, proximity, activity/motion, engagement) of persons **14** in crowd/group level interactions.

This written description uses examples to disclose the embodiments, including the best mode, and also to enable any person skilled in the art to practice the disclosed subject matter, including making and using any devices or systems and performing any incorporated methods. The patentable scope of the subject matter is defined by the claims, and may include other examples that occur to those skilled in the art. Such other examples are intended to be within the scope of the claims if they have structural elements that do not differ from the literal language of the claims, or if they include equivalent structural elements with insubstantial differences from the literal languages of the claims.

The invention claimed is:

1. A method, comprising:

receiving data from one or more cameras in an environment;

generating metadata of one or more video analytics streams produced from the data;

generating one or more time series of values based on the metadata;

generating one or more affect scores for the one or more time series;

generating a first signature representative of an observed behavior based on the one or more affect scores;

performing pairwise matching by determining whether the first signature matches a second signature indicative of a query behavior; and

performing an action when the first signature matches the second signature.

2. The method of claim **1**, wherein the pairwise matching is performed after only a single observation of an instance of the query behavior is obtained.

3. The method of claim **1**, wherein pairwise matching comprises deriving pairwise match scores in accordance with the following representation:

$$d(b_k, b_j) = \sum_i^n w_i \| sg_i(b_k) - sg_i(b_j) \|$$

where d is a distance measure, b is an observed behavior, sg is a signal generator, n is a number of signal generators, and w is a weight associated with each signal generator.

4. The method of claim **1**, wherein the one or more cameras comprise red, green, blue, depth (RGB+D) cameras that capture estimates of location and articulated body motion, and fixed cameras and pan tilt zoom (PTZ) cameras that capture facial imagery.

5. The method of claim **1**, wherein the video analytics stream comprises a set of person descriptors that encode locations of individuals in site coordinates, motion signatures of the individuals, expression profiles of the individuals, gaze direction of the individuals, or some combination thereof.

6. The method of claim **1**, wherein the video analytics stream is produced by:

tracking individuals via the one or more cameras,

generating a motion signature for each individual based on space-time interest points;

capturing facial images using the one or more cameras; and

estimating facial expression and gaze direction based on the facial images.

7. The method of claim **1**, wherein the values in the time series range from 0 to 1.

8. The method of claim **1**, wherein the one or more affect scores range from 0 to 1.

9. The method of claim **1**, wherein the generation of the metadata, the one or more time series of values, the one or more affect scores, and the signature are performed by a signal generator bank module.

10. The method of claim **1**, wherein performing the action comprising sounding an alarm, calling emergency services, triggering an alert, sending a message, displaying an alert, or some combination thereof when the first signature matches the second signature.

11. The method of claim **1**, comprising determining weights used to generate the one or more affect scores by performing machine learning on a training set of pairs of behaviors that are labeled as positive matches and pairs of behaviors that are labeled as negative matches.

12. One or more tangible, non-transitory computer-readable media storing computer instructions that, when executed by one or more processors, cause the one or more processors to:

receive data from one or more cameras in an environment;

generate metadata of one or more video analytics streams produced from the data;

generate one or more time series of values based on the metadata;

generate one or more affect scores for the one or more time series;

generate a first signature representative of an observed behavior based on the one or more affect scores;

perform pairwise matching by determining whether the first signature matches a second signature indicative of a query behavior; and

provide an output when the first signature matches the second signature indicative of the query behavior.

13. The one or more computer-readable media of claim **12**, wherein the pairwise matching is performed after only a single observation of an instance of the query behavior is obtained.

14. The one or more computer-readable media of claim **12**, wherein the computer instructions cause the one or more processors to produce the video analytics stream by:

tracking individuals via the one or more cameras,

generating a motion signature for each individual based on space-time interest points;

capturing facial images using the one or more cameras; and

estimating facial expression and gaze direction based on the facial images.

15. The one or more computer-readable media of claim **12**, wherein pairwise matching comprises deriving pairwise match scores in accordance with the following representation:

$$d(b_k, b_j) = \sum_i^n w_i \| sg_i(b_k) - sg_i(b_j) \|$$